ORGANIC LETTERS

2005
Vol. 7, No. 26
5873−5876

Enzymatic Formation of Indole-Containing Unnatural Cyclic Polyprenoids by Bacterial Squalene:Hopene Cyclase

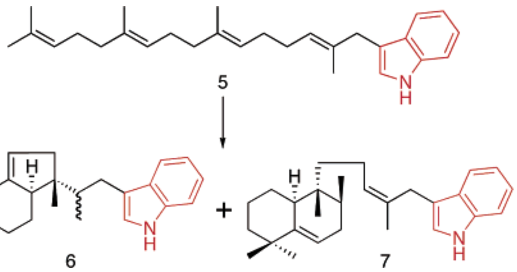
Hideya Tanaka,[†] Hiroshi Noguchi,[†] and Ikuro Abe*[,†,¶]

School of Pharmaceutical Sciences and the COE 21 Program, University of Shizuoka, Shizuoka 422-8526, Japan, and PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

abei@ys7.u-shizuoka-ken.ac.jp

Received October 17, 2005

ABSTRACT

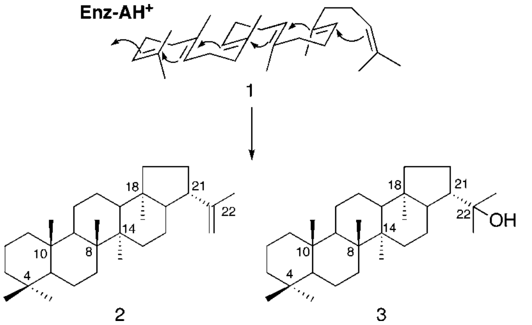


Two indole-containing substrate analogues, in which a C_{20} isoprene unit is connected to indole (3-(geranylgeranyl)indole and 3-(farnesyldimethylallyl)indole), were synthesized and tested for enzymatic cyclization by squalene:hopene cyclase from *Alicyclobacillus acidocaldarius.* Interestingly, 3-(geranylgeranyl)indole was not a substrate for the bacterial squalene cyclase, while 3-(farnesyldimethylallyl)indole was efficiently converted to a 2:1 mixture of unnatural novel products.

Squalene:hopene cyclase (SHC) (E.C. 5.4.99.7) from a thermoacidophilic bacteria, *Alicyclobacillus acidocaldarius*, catalyzes remarkable cyclization of squalene (**1**) into a 5:1 mixture of hop-22(29)-ene (**2**) and hopan-22-ol (**3**) (Scheme 1). The enzyme binds the substrate in an all-*chair* conformation and mediates sequential C−C bond formation reactions through a progression of rigidly held carbocationic intermediates. Recent crystallographic and structure-based mutagenesis studies as well as utilization of active-site probes have revealed intimate structural details of the enzyme-templated polyene cyclization reaction.[1,2]

Interestingly, the bacterial SHC exhibits remarkably broad substrate tolerance. The enzyme accepts a wide variety of nonphysiological substrate analogues (C_{15}−C_{35}) and efficiently performs sequential ring-forming reactions to produce a series of unnatural cyclic polyprenoids.[3] On the basis of crystal structure of the enzyme, the hydrophobic

Scheme 1. Enzymatic Cyclization of Squalene by SHC



[†] University of Shizuoka.
[¶] PRESTO.
(1) For reviews, see: (a) Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev.* **1993**, *93*, 2189−2206. (b) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 2812−2833. (c) Hoshino, T.; Sato, T. *Chem. Commun.* **2002**, 291−301.

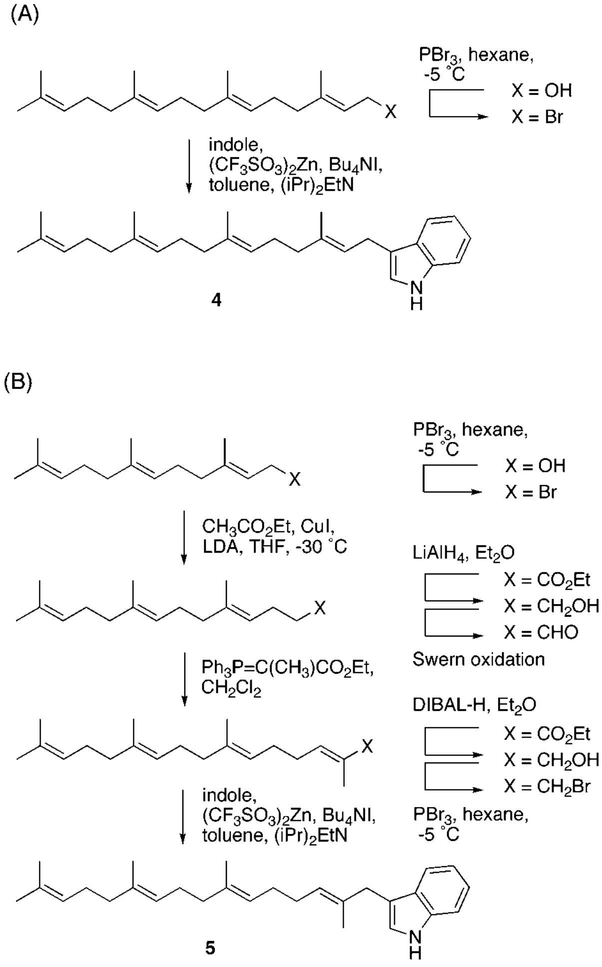
10.1021/ol052507q CCC: $30.25

active-site cavity lined with aromatic amino acid residues appears to have enough space to accept larger substrate analogues.[2] One of the most impressive examples is the enzymatic formation of a supranatural hexacyclic polyprenoid with a 6/6/6/6/6/5-fused ring system from a C_{35} analogue in which a farnesyl C_{15} unit is connected in a head-to-head fashion to a geranylgeranyl C_{20} unit.[3a] It is remarkable that even for the unnatural substrate with an additional isoprene unit stereochemistry of the cyclization reaction was strictly controlled by the enzyme, leading to formation of 11 chiral centers in a regio- and stereospecific manner. Manipulation of the enzyme reaction by substrate analogues would thus lead to further production of chemically and structurally disparate unnatural polycyclic polyprenoids.

We report herein the synthesis and enzymatic cyclization of two indole-containing substrate analogues in which a C_{20} isoprene unit is connected to indole, 3-(geranylgeranyl)indole (**4**)[4] and 3-(farnesyldimethylallyl)indole (**5**).[5] Substrate **4** is a putative precursor for natural indole diterpenes,[6a,7] and enzymatic cyclization of 3-(ω-oxidogeranylgeranyl)indole into a hexacyclic petromindole by a plant oxidosqualene:lupeol cyclase from *Arabidopsis thaliana* has been recently reported by Matsuda and co-workers.[6a] Substrate **5** has a farnesyl C_{15} unit as in the case of the natural substrate, squalene (C_{15} unit $\times$ 2). Our previous studies have suggested that the presence of the farnesyl C_{15} unit in the substrate is important for enzymatic formation of polycyclic products by *A. acidocaldarius* SHC.[3a,c] Further, the $\pi-\pi$ interactions of the indole ring with the active-site aromatic residues of the enzyme were also anticipated for both substrate analogues.

The convergent synthesis of **4** involved regioselective alkylation of indole at C3 by geranylgeranyl C_{20} bromide mediated by zinc triflate according to the literature (Scheme 2A).[6] On the other hand, **5**, in which a farnesyl C_{15} unit is connected in a head-to-head fashion to a dimethylallyl C_5 unit, was synthesized starting from farnesol,[8] and finally coupled with indole in the same manner (Scheme 2B).

Scheme 2. Synthesis of Indole-Containing Substrate Analogues



When incubated with purified recombinant *A. acidocaldarius* SHC,[9] 3-(geranylgeranyl)indole (**4**) did not afford any cyclization product (Scheme 3A), which was confirmed by TLC and GLC analysis. This is in sharp contrast with the above-mentioned enzymatic cyclization of 3-(ω-oxidogeranylgeranyl)indole by the plant oxidosqualene cyclase.[6a] This

---

(2) For crystal structure of *A. acidocaldarius* SHC, see: (a) Wendt, K. U.; Poralla, K.; Schulz, G. E. *Science* **1997**, *277*, 1811−1815. (b) Wendt, K. U.; Lenhart, A.; Schulz, G. E. *J. Mol. Biol.* **1999**, *286*, 175−187.

(3) For enzymatic conversion of squalene analogues ($C_{15}-C_{35}$) by bacterial squalene cyclase, see: (a) Abe, I.; Tanaka, H.; Noguchi, H. *J. Am. Chem. Soc.* **2002**, *124*, 14514−14515. (b) Tanaka, H.; Noguchi, H.; Abe, I. *Org Lett.* **2004**, *6*, 803−806. (c) Tanaka, H.; Noguchi, H.; Abe, I. *Tetrahedron Lett.* **2004**, *45*, 3093−3096. (d) Rohmer, M.; Anding, C.; Ourisson, G. *Eur. J. Biochem.* **1980**, *112*, 541−547. (e) Bouvier, P.; Berger, Y.; Rohmer, M.; Ourisson, G. *Eur. J. Biochem.* **1980**, *112*, 549−556. (f) Rohmer, M.; Bouvier, P.; Ourisson, G. *Eur. J. Biochem.* **1980**, *112*, 557−560. (g) Renoux, J.-M.; Rohmer, M. *Eur. J. Biochem.* **1986**, *155*, 125−132. (h) Abe, I.; Rohmer, M. *J. Chem. Soc., Perkin Trans. 1* **1994**, 783−791. (i) Abe, I.; Dang, T.; Zheng, Y. F.; Madden, B. A.; Feil, C.; Poralla, K.; Prestwich, G. D. *J. Am. Chem. Soc.* **1997**, *119*, 11333−11334. (j) Robustell, B.; Abe, I.; Prestwich, G. D. *Tetrahedron Lett.* **1998**, *39*, 957−960. (k) Robustell, B.; Abe, I.; Prestwich, G. D. *Tetrahedron Lett.* **1998**, *39*, 9385−9388. (l) Zheng, Y. F.; Abe, I.; Prestwich, G. D. *J. Org. Chem.* **1998**, *63*, 4872−4873. (m) Sato, T.; Abe, T.; Hoshino, T. *Chem. Commun.* **1998**, 2617−2618. (n) Hoshino, T.; Kondo, T. *Chem. Commun.* **1999**, 731−732. (o) Hoshino, T.; Ohashi, S. *Org. Lett.* **2002**, *4*, 2553−2556. (p) Hoshino, T.; Nakano, S.; Kondo, T.; Sato, T.; Miyoshi, A. *Org. Biomol. Chem.* **2004**, *2*, 1456−1470. (q) Nakano, S.; Ohashi, S.; Hoshino, T. *Org. Biomol. Chem.* **2004**, *2*, 2012−2022. (r) Hoshino, T.; Kumai, Y.; Kudo, I.; Nakano, S.; Ohashi, S. *Org. Biomol. Chem.* **2004**, *2*, 2650−2657.

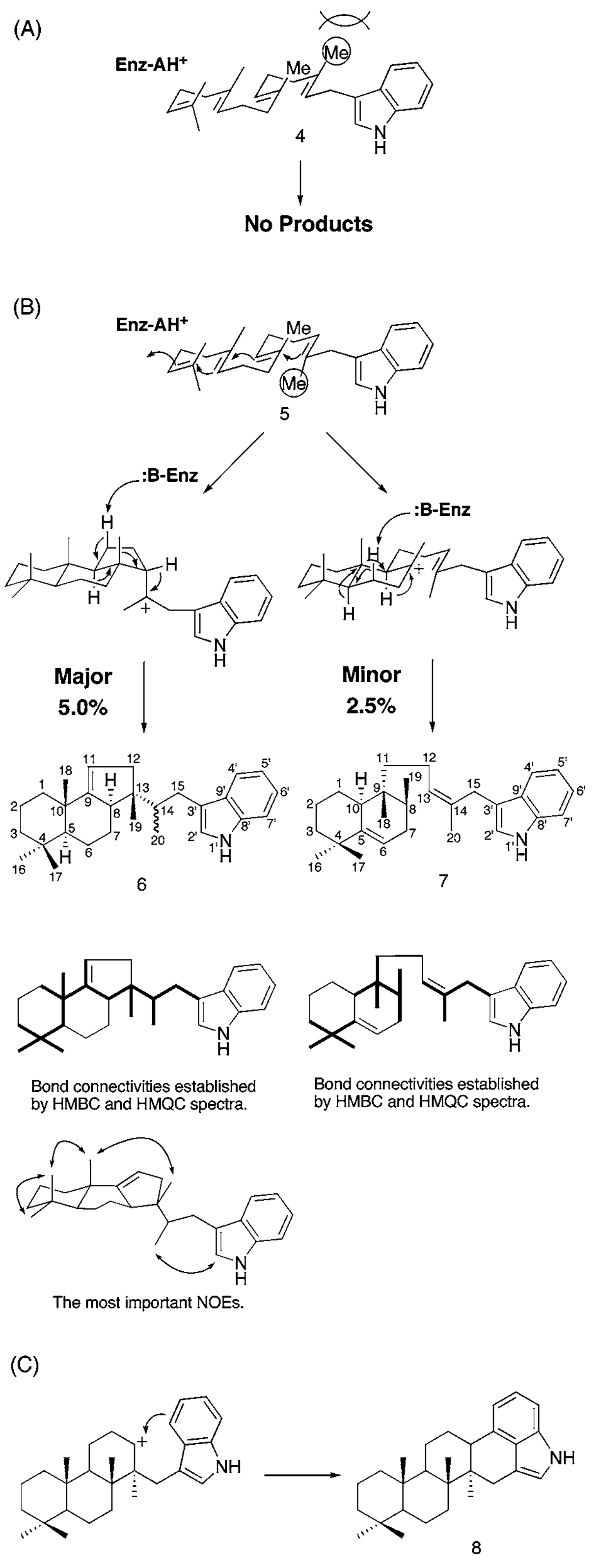
(4) 3-((2*E*,6*E*,10*E*)-3,7,11,15-tetramethylhexadeca-2,6,10,14-tetraenyl)indole (**4**): ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.87 (brs, 1H), 7.61 (dtd, 1H, $J = 7.8$, 1.1, 0.9 Hz), 7.35 (dt, 1H, $J = 8.3$, 0.9 Hz), 7.20 (ddd, 1H, $J = 8.2$, 7.0, 1.1 Hz), 7.13 (ddd, 1H, $J = 7.9$, 7.0, 1.1 Hz), 6.95 (dt, 1H, $J = 2.3$, 1.0 Hz), 5.48 (dd, 1H, $J = 7.8$, 6.8 Hz), 5.28 (m, 3H), 3.48 (d, 2H, $J = 6.8$ Hz), 2.09 (m, 12H), 1.78 (s, 3H, Me-17), 1.70 (s, 3H, Me-18), 1.62 (s, 6H, Me-19, 20), 1.61 (s, 3H, Me-21). ^{13}C NMR (100 MHz, CDCl$_3$): δ 136.5, 135.6, 135.0, 134.9, 131.2, 127.5, 124.4, 124.3, 123.6, 122.9, 121.9, 121.1, 119.1, 119.0, 116.1, 111.0, 39.7, 26.8, 26.7, 26.6, 25.7, 24.0, 17.7, 16.1, 16.0, 16.0. HRMS (FAB): found for $[C_{28}H_{39}N]^+$ 389.3059; calcd 389.3082.

(5) 3-((2*E*,6*E*,10*E*)-2,7,11,15-tetramethylhexadeca-2,6,10,14-tetraenyl)indole (**5**): ^{1}H NMR (400 MHz, CDCl$_3$): δ 7.91 (brs, 1H), 7.62 (ddt, 1H, $J = 7.8$, 1.1, 0.9 Hz), 7.34 (dt, 1H, $J = 8.3$, 0.9 Hz), 7.18 (ddd, 1H, $J = 8.2$, 7.0, 1.2 Hz), 7.10 (ddd, 1H, $J = 7.9$, 7.0, 1.1 Hz), 6.97 (dt, 1H, $J = 2.3$, 1.1 Hz), 5.39 (m, 1H), 5.14 (m, 3H), 3.46 (s, 2H), 2.04 (m, 12H), 1.70 (s, 3H, Me-17), 1.63 (s, 3H, Me-18), 1.62 (s, 9H, Me-19, 20, 21). ^{13}C NMR (100 MHz, CDCl$_3$): δ 136.4, 135.1, 134.9, 134.2, 131.2, 127.9, 125.4, 124.4, 124.3, 122.0, 121.8, 119.4, 119.1, 114.8, 110.9, 39.8, 39.7, 35.8, 28.4, 28.2, 26.8, 26.7, 25.7, 17.7, 16.1, 16.0. HRMS (FAB): found for $[C_{28}H_{39}N]^+$ 389.3094; calcd 389.3082.

(6) (a) Xiong, Q.; Zhu, X.; Wilson, W. K.; Ganesan, A.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2003**, *125*, 9002−9003. (b) Zhu, X.; Ganesan, A. *J. Org. Chem.* **2002**, *67*, 2705−2708.

(7) Fueki, S.; Tokiwano, T.; Toshima, H.; Oikawa, H. *Org. Lett.* **2004**, *6*, 2697−2700.

Scheme 3. Enzymatic Conversion of Indole-Containing Analogues



suggested that the bacterial SHC is particularly sensitive to the structural changes on the pro-C14β face (numbering in hopene) (Scheme 3A) and thus fails to bind the substrate **4**. As we have previously proposed, α-orientation of the *pro*-C14 methyl group is crucial to the correct folding and binding of the substrate, as in the case of the cyclization of squalene into hopene.[3a,c] Presumably, the alternative β-orientation of the *pro*-C14 methyl group may interact repulsively with the *pro*-C8 methyl or a nearest neighbor from the substrate binding pocket of the enzyme.[3a]

In contrast, 3-(farnesyldimethylallyl)indole (**5**) was a competent substrate for the enzyme and efficiently converted to a 2:1 mixture of a tetracyclic and a pentacyclic product, **6** and **7**, which were completely separated by reverse-phase HPLC (Scheme 3B).[9] The structures of the cyclization products were determined by NMR (^{1}H and ^{13}C NMR, HMQC, HMBC, NOESY, and differential NOE) and MS spectroscopy.[10,11] The ^{1}H NMR spectrum of the major product **6** (1.0 mg, 5.0% yield) revealed the presence of four methyl singlets (δ 0.88, 0.88, 0.92, and 1.10), one methyl doublet (δ 0.78, $J = 6.9$ Hz), and one vinylic proton (δ 5.19), in addition to the signals due to the indole ring.[10]

A structure with a 6/6/5-fused tricyclic ring moiety was uniquely consistent with both biogenetic reasoning and the heteronuclear correlation spectroscopy (HMQC and HMBC). Further, the ring junctions and the stereochemistry of ring substituents were established by NOE experiments. Interestingly, formation of a similar 6/6/5-fused tricyclic ring moiety with the $\Delta^{9(11)}$ double bond has been recently reported for one of the enzymatic cyclization products of (3*S*)-2,3-

---

(8) (a) Coates, R. M.; Ley, D. A.; Cavender, P. L. *J. Org. Chem.* **1978**, *43*, 4915−4922. (b) Zheng, Y. F.; Oehlschlager, A. C. *J. Org. Chem.* **1994**, *59*, 5803−5809.

(9) The recombinant *A. acidocaldarius* SHC was prepared as described in the previous papers.[3a−c] The reaction mixture contained the substrate analogue **5** (or **4**) (20 mg) and purified recombinant SHC (120 mg) in 400 mL of 50 mM Na citrate, pH 6.0, 0.1% Triton X-100, and was incubated at 60 °C for 16 h. The incubations were stopped by freezing and lyophilization, followed by extraction with 200 mL of hexane (×3). The combined extracts were evaporated to dryness, separated on SiO$_2$ TLC (20% EtOAc in hexane) and by reverse-phase HPLC (Phenomex Gemini 5m ODS column, 250 × 4.6 mm, 5% THF in MeCN, 0.5 mL/min) to give 1.0 mg of **6** (*R*t = 21.9 min) and 0.5 mg of **7** (*R*t = 20.3 min) (from **5**).

(10) Major product (**6**): ^{1}H NMR (500 MHz, CDCl$_3$): δ 7.92 (brs, 1H, NH), 7.64 (dt, 1H, $J = 7.9$, 1.1 Hz), 7.36 (dt, 1H, $J = 8.0$, 1.0 Hz), 7.19 (ddd, 1H, $J = 8.0$, 7.0, 1.1 Hz), 7.12 (ddd, 1H, $J = 8.0$, 7.0, 1.0 Hz), 6.98 (d, 1H, $J = 1.7$ Hz), 5.19 (m, 1H), 3.07 (dd, 1H, $J = 14.0$, 2.6 Hz), 2.77 (brd, 1H), 2.36 (1H), 2.36 (1H), 1.10 (s, 3H, Me-18), 0.92 (s, 3H, Me-17), 0.88 (s, 3H, Me-16), 0.88 (s, 3H, Me-19), 0.78 (d, 3H, $J = 6.9$ Hz, Me-20). ^{13}C NMR (125 MHz, CDCl$_3$): δ 158.4 (C-9), 136.4 (C-8′), 127.8 (C-9′), 121.7 (C-2′), 121.7 (C-6′), 119.1 (C-5′), 119.0 (C-4′), 116.4 (C-3′), 115.5 (C-11), 111.0 (C-7′), 50.26 (C-8), 48.2 (C-13), 47.0 (C-12), 45.7 (C-5), 44.6 (C-14), 42.6 (C-3), 40.1 (C-1), 36.4 (C-10), 33.4 (C-4), 33.0 (C-16), 28.4 (C-15), 25.1 (C-18), 22.1 (C-7), 21.5 (C-17), 19.4 (C-5), 19.2 (C-2), 17.0 (C-19), 15.7 (C-20). The NMR assignments were performed according to data from H−H COSY, HMQC, HMBC, NOESY, and differential NOE experiments. LRMS (FAB$^+$, %): *m*/*z* 130 (98), 158 (67), 231 (100), 390 (M$^+$, 15). HRMS (FAB$^+$): found for [C$_{28}$H$_{39}$N] 389.3059; calcd 389.3082. $[\alpha]^{25}_{D} = -6.7°$ ($c = 0.10$ in CHCl$_3$).

(11) Minor product (**7**): ^{1}H NMR (500 MHz, CDCl$_3$): δ 7.93 (brs, 1H, NH), 7.62 (dt, 1H, $J = 8.0$, 1.1 Hz), 7.35 (dt, 1H, $J = 8.2$, 1.0 Hz), 7.17 (ddd, 1H, $J = 8.2$, 7.0, 1.2 Hz), 7.09 (ddd, 3H, $J = 8.0$, 7.0, 1.2 Hz), 6.97 (t, 1H, $J = 1.1$ Hz), 5.43 (m, 1H), 5.34 (m, 1H), 3.44 (s, 2H, Me-30), 1.95 (m, 2H), 1.63 (s, 3H, Me-34), 1.55 (s, 3H, Me-34), 1.06 (s, 3H, Me-17), 1.00 (s, 3H, Me-16), 0.83 (d, 3H, $J = 6.8$ Hz, Me-19), 0.61 (s, 3H, Me-18). ^{13}C NMR (125 MHz, CDCl$_3$): δ 146.3 (C-5), 133.6 (C-14), 133.6 (C-8′), 126.1 (C-13), 126.0 (C-9′), 122.1 (C-9), 122.1 (C-6′), 121.8 (C-2′), 119.4 (C-4′), 119.1 (C-5′), 116.1 (C-6), 114.9 (C-3′), 110.9 (C-7′), 41.0 (C-3), 39.9 (C-10), 39.9 (C-10), 37.1 (C-15), 36.7 (C-11), 35.8 (C-4), 35.8 (C-9), 35.8 (C-15), 33.5 (C-8), 31.7 (C-7), 29.8 (C-17), 29.0 (C-16), 27.6 (C-1), 22.3 (C-2), 21.5 (C-12), 16.1 (C-18), 15.9 (C-20), 15.2 (C-19). The NMR assignments were performed according to data from H−H COSY, HMQC, HMBC, and differential NOE experiments. LRMS (FAB$^+$, %): *m*/*z* 130 (100), 184 (48), 390 (M$^+$, 30). HRMS (FAB$^+$): found for [C$_{28}$H$_{39}$N] 389.3059; calcd 389.3082. $[\alpha]^{25}_{D} = 0.1°$ ($c = 0.10$ in CHCl$_3$).

oxidosqualene by a Gly600 deletion mutant of *A. acidocaldarius* SHC.[12a]

On the other hand, the ^{1}H NMR spectrum of the minor product **7** (0.5 mg, 2.5% yield) indicated the presence of doublet (δ 0.83, $J = 6.8$ Hz), one vinylic methyl (δ 1.63), three methyl singlets (δ 0.61, 1.00, and 1.06), one methyl, and one vinylic proton (δ 5.43),[11] which is in good agreement with those of a bicyclic product from squalene by a F365A mutant of *A. acidocaldarius* SHC.[12b] Indeed, NMR spectroscopic data (HMQC, HMBC, and NOE) unambiguously established the structure with a 6/6-fused bicyclic ring moiety with the Δ^5 double bond.

Enzymatic formation of a petromindole-like hexacyclic product **8** from 3-(farnesyldimethylallyl)indole (**5**) would apparently involve the formation of a tricyclic secondary cationic intermediate with a 6/6/6-fused ring system and a subsequent ring closure to indole at C4′ (Scheme 3C). However, the cyclization reaction of **5** by *A. acidocaldarius* SHC, initiated in a *chair*−*chair*−*chair* conformation, was interrupted at the bicyclic or tricyclic cationic stage probably because of the presence of the bulky indole ring. In addition, the $\pi-\pi$ interactions of the indole ring with the active-site aromatic residues (e.g., Trp169 and Phe605 on the basis of the crystal structure of *A. acidocaldarius* SHC[2]) would also be anticipated, which may affect the folding conformation of the substrate at the active site of the enzyme as well as the subsequent cyclization and rearrangement reactions. As a result, from the tricyclic Markovnikov cation with a 6/6/5-fused ring system, a backbone rearrangement (H-13β→14, CH$_3$-8β→13β, H-9α→8α) with elimination of H-11β yielded the unnatural novel pentacyclic **6** as the major product, while from the bicyclic tertiary cation with a 6/6-fused ring system, a backbone rearrangement (H-9α→8α, CH$_3$-10β→9β, H-5α→10α) with elimination of H-6β produced the unnatural novel tetracyclic **7** as the minor product (Scheme 3B). Remarkably, the enzyme strictly controlled the stereochemistry of the cyclization and the rather unusual rearrangement reactions even for the indole-containing substrate analogue. Further, it is also interesting that triterpenes with the similar tricyclic and bicyclic skeletons have been reported for the products of *A. acidocaldarius* SHC mutants as mentioned above.[12]

For further production of structurally divergent unnatural natural products, manipulation of the enzyme reaction by substrate analogues with different isoprene chain length and/or with other heteroaromatic ring systems is now in progress in our laboratories.

Acknowledgment. This work was in part supported by the COE21 Program, Grant-in-Aid for Scientific Research (Nos. 16510164, 17310130, and 17035069), and Cooperation of Innovative Technology and Advanced Research in Evolutional Area (CITY AREA) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. H.T. is a recipient of the JSPS Fellowship for Young Scientist (No. 175479).

Supporting Information Available: Complete set of spectroscopic data (^{1}H and ^{13}C NMR, HMQC, HMBC, and NOE) of the substrate analogues (**4**, **5**) and enzyme reaction products (**6**, **7**) (15 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

OL052507Q

---

(12) (a) Hoshino, T.; Shimizu, K.; Sato, T. *Angew. Chem., Int. Ed.* **2004**, *43*, 6700−6703. (b) Hoshino, T.; Sato, T. *Chem. Commun.* **1999**, 2005−2006.